Amendments to the Claims:

This listing of claims will replace all prior versions, and listings of claims in the application:

Listing of Claims:

1. (Previously Presented) A method for processing query messages over a network, the method comprising:

extracting a plurality of queries from a plurality of query messages received from a plurality of users over the network;

creating a first request message including the plurality of queries and a first sequence number associated with one or more of the queries;

sending the first request message to a search engine;

receiving a response message from the search engine, the response message including a plurality of replies and the first sequence number, wherein the first sequence number is associated with one or more of the replies, and wherein each reply associated with the first sequence number is generated in response to a query also associated with the first sequence number;

creating a plurality of reply messages from the plurality of replies; and

sending the plurality of reply messages to the plurality of users over the network.

2-3. (Canceled).

4. (Original) The method of claim 1, further comprising:

determining a message latency associated with the first sequence number.

5. (Original) The method of claim 4, wherein said determining a message latency includes:

updating a request timestamp based on the request message;

updating a response timestamp based on the response message; and

comparing the request timestamp and the response timestamp.

6. (Previously Presented) The method of claim 5, further comprising:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies; and

updating the response timestamp based on the additional response message.

7. (Original) The method of claim 4, wherein said determining a message latency includes:

updating a query count based on the request message;

updating a reply count based on the response message; and

comparing the query count and the reply count.

8. (Previously Presented) The method of claim 7, wherein said determining a message latency includes:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies; and

updating the reply count based on the additional response message.

9. (Original) The method of claim 4, wherein said determining a message latency includes:

updating a response count based on the response message; and

comparing the response count to a predetermined response count.

10. (Previously Presented) The method of claim 9, wherein said determining a message latency includes:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies; and

updating a response count based on the additional response message.

11. (Currently Amended) A system for processing query messages over a network, comprising:

a first network interface coupled to a first network;

a second network interface coupled to a second network;

at least one processor coupled to the first network interface and the second network interface; and

a memory coupled to the processor, the memory including instructions adapted to be executed by the processor to:

extract a plurality of queries from a plurality of query messages received from a plurality of users over the first network interface;

create a first request message including the plurality of queries and a first sequence number associated with one or more of the queries;

send the first request message to a search engine over the second network interface;

receive a response message from the search engine, the response message including a plurality of replies~~,~~ and the first sequence number, wherein the first sequence number is associated with one or more of the replies, and

wherein each reply associated with the first sequence number is generated in response to a query also associated with the first sequence number ;

create a plurality of reply messages from the plurality of replies; and

send the plurality of reply messages to the plurality of users over the first network interface.

12. (Original) The system of claim 11, wherein the first network and the second network are the same network.

13-14. (Canceled).

15. (Previously Presented) The system of claim 11, wherein the instructions are further adapted to:

determine a message latency associated with the first sequence number, including:

update a request timestamp based on the request message;

update a response timestamp based on the response message; and

compare the request timestamp and the response timestamp;

receive an additional response message from the search engine, the additional response message including an additional plurality of replies; and

update the request timestamp based on the additional response message.

16. (Previously Presented) The system of claim 11, wherein the instructions are further adapted to:

determine a message latency associated with the first sequence number, including:

update a query count based on the request message;

update a reply count based on the response message; and

compare the query count and the reply count;

receive an additional response message from the search engine, the additional response message including an additional plurality of replies; and

update the reply count based on the additional response message.

17. (Previously Presented) The system of claim 11, wherein the instructions are further adapted to:

determine a message latency associated with the first sequence number, including:

update a response count based on the response message; and

compare the response count to a predetermined response count;

receive an additional response message from the search engine, the additional response message including an additional plurality of replies; and

update a response count based on the additional response message.

18. (Previously Presented) A computer readable medium storing instructions adapted to be executed by at least one processor to implement a method for processing query messages over a network, the method comprising:

extracting a plurality of queries from a plurality of query messages received from a plurality of users over the network;

creating a first request message including the plurality of queries and a first sequence number associated with one or more of the queries;

sending the first request message to a search engine;

receiving a response message from the search engine, the response message including a plurality of replies~~,~~ and the first sequence number, wherein the first sequence number is associated with one or more of the replies, and

wherein each reply associated with the first sequence number is generated in response to a query also associated with the first sequence number;

creating a plurality of reply messages from the plurality of replies; and

sending the plurality of reply messages to the plurality of users over the network.

19-20. (Canceled).

21. (Original) The computer readable medium of claim 18, wherein the method further comprises:

determining a message latency associated with the first sequence number.

22. (Original) The computer readable medium of claim 21, wherein said determining a message latency includes:

updating a request timestamp based on the request message;

updating a response timestamp based on the response message; and

comparing the request timestamp and the response timestamp.

23. (Previously Presented) The computer readable medium of claim 22, wherein the method further comprises:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies; and

updating the request timestamp based on the additional request message.

24. (Original) The computer readable medium of claim 21, wherein said determining a message latency includes:

updating a query count based on the request message;

updating a reply count based on the response message; and

comparing the query count and the reply count.

25. (Previously Presented) The computer readable medium of claim 24, wherein said determining a message latency includes:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies; and

updating the reply count based on the additional response message.

26. (Original) The computer readable medium of claim 21, wherein said determining a message latency includes:

updating a response count based on the response message.

27. (Previously Presented) The computer readable medium of claim 26, wherein said determining a message latency includes:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies; and

updating a response count based on the additional response message.

28. (Previously Presented) The method of claim 1, wherein the response message includes one or more replies generated in response to the first request message and one or more replies generated in response to a second request message, wherein the second request message is created after the first request message.

29. (Previously Presented) The method of claim 1, wherein the first request message includes state information or a reference to the state information for each query, and wherein the response message includes state information or a reference to the state information for each reply, wherein sending the plurality of reply messages to the plurality of users comprises identifying a user associated with each query from which each reply message was generated using the state information.

30. (Previously Presented) The method of claim 1, wherein the first sequence number uniquely identifies the one or more of the queries.

31-32. (Canceled).

33. (Previously Presented) The method of claim 1, wherein the response message further includes a second sequence number that is associated with one or more replies that are not associated with the first sequence number.

34. (Previously Presented) The method of claim 1, wherein each query message is a request to resolve a domain name.

35. (Previously Presented) The method of claim 34, wherein extracting the plurality of queries from the plurality of query messages is performed by a front-end protocol engine that sends the request message via a wide area network to the search engine.